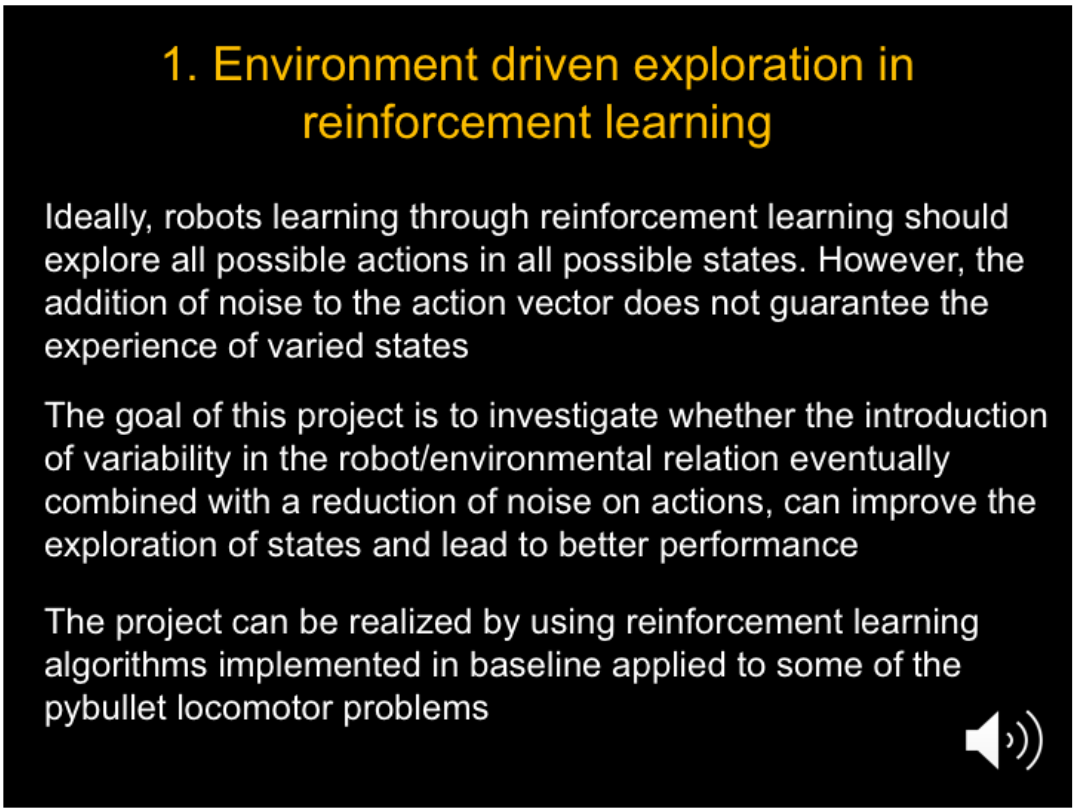
In this presentation I will illustrate 11 suggested research projects



The first project aims to investigate environment driven exploration in reinforcement learning

As I explained in Lecture 6, reinforcement learning algorithms rely on stochastic policy to explore variations of the current policy. In policy learning methods variations are normally introduced by executing an actions constituted by perturbed version of the output vectors.

IN PRINCIPLE, the identification of the optimal policy requires that the robot try every possible action in every possible state. However, the addition of noise in the action vector does no guarantee that the robot will experience all possible state. Moreover, the usage of stochastic policies prevent the possibility to discover precise strategies, i.e. strategies in which the execution of specific action is necessary to achieve the goal and in which the execution of randomly perturbed version of those actions does not permit the achievement of the goal.

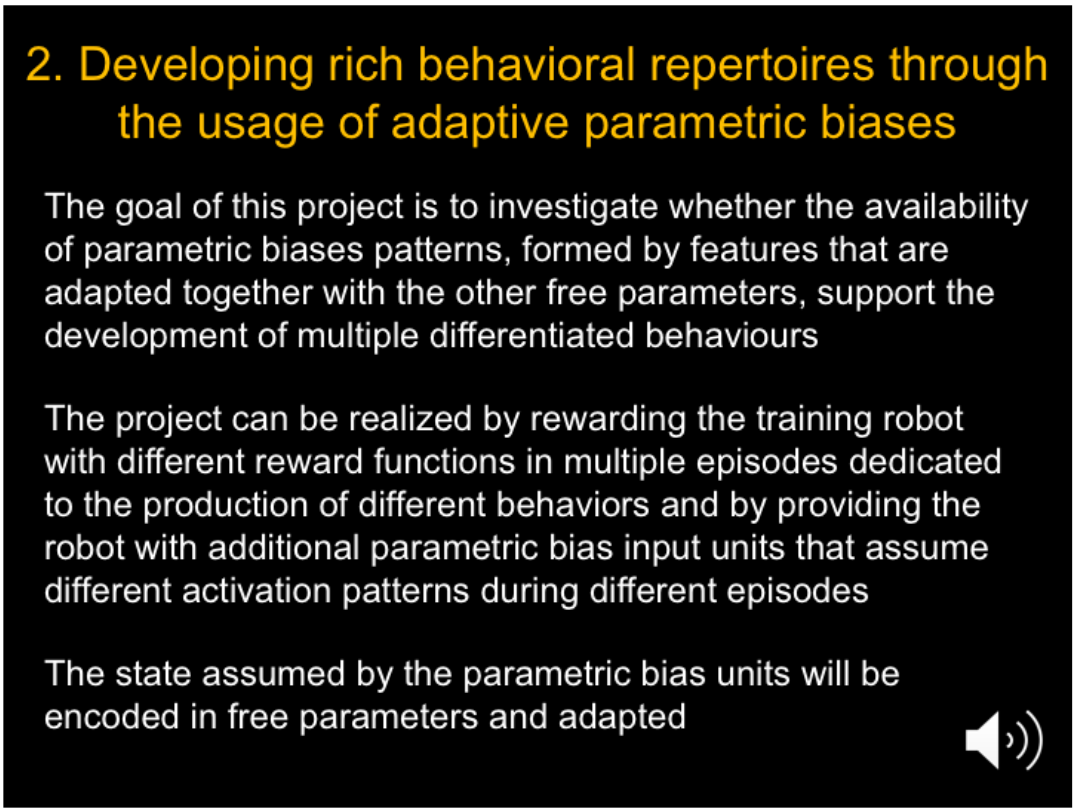
An alternative way to introduce variations consists in introducing variations in the environment or in the agent/environmental relations. Variations of this type enable the robot to experience varied states. This method is normally used to introduce some limited variation at the beginning of each episode

which however do not ensure that the robot keep experience sufficient variations later on.

THE GOAL of this project is that to investigate whether the introduction of a greater amount of environmental variations, eventually combined with a reduction of the amount of variations introduced in the actions executed by the robot, can produce better results.

Environmental variations can be introduced, for example, by adding random force to the body parts of the robot that alter the posture of the robot during motion and/ or by the introduction variation in the surface over which the robot is situated.

The PROJECT can be realized by using reinforcement learning algorithms implemented in baseline applied to some of the pybullet locomotor problems



This second project address the development of rich behavioral repertoires through the usage of adaptive parametric biases

Adaptive methods are typically applied to the synthesis of a single behavioral capability, e.g. walking at the fastest speed or navigating toward a certain destination. The way in which these methods can be scaled to robots capable to display a reach behavioral repertoire constitutes a open issue.

The GOAL of this project is to investigate whether the availability of adaptable parametric biases can support the development of multiple behaviors.

Parametric biases (see Tani J. & Masato I. (2003). Self-Organization of Behavioral Primitives as Multiple Attractor Dynamics, IEEE Transactions on Systems, Man and Cybernetics: Part A, 33 (4): 481-488.) consist of perceptual patterns provided by the experimenter that differentiate the contexts that require the exhibition of different behaviors.

Parametric biases thus play the role of affordances, i.e. of perceptual state affording the execution of an associated behaviour. An example of

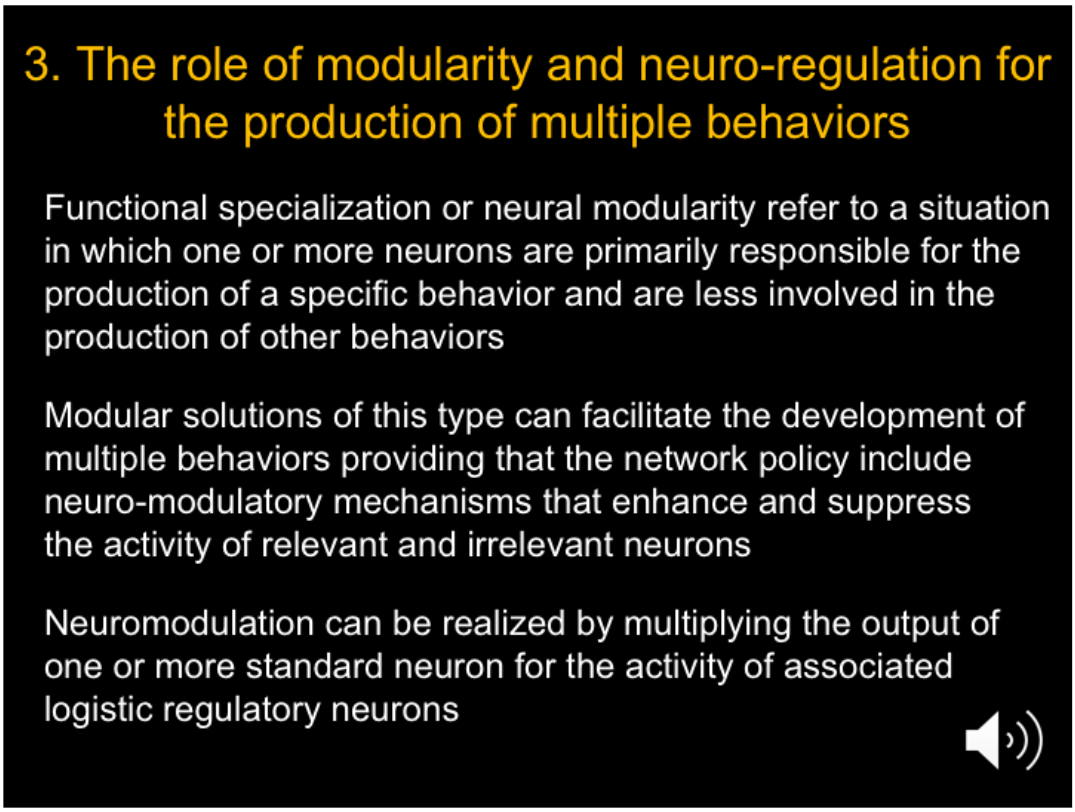
affordance in the case of a soccer playing robot is constituted by the perception of the ball in the proximal space of the robot accompanied by the perception of the goal within a given maximum distance. The perception of this state afford the execution of a shot on goal behaviour.

The features that form each parametric bias vector can be specified by the experimenter manually or can be adapted together with the other free parameters of the policy. This latter possibility permit to shape the features of the parametric bias vectors in a way that facilitate the execution of the corresponding behaviour.

The PROJECT can be realized by rewarding the training robot with different reward functions in multiple episodes dedicated to the production of different behaviors and by providing the robot with additional parametric bias input units that assume different activation patterns during different episodes.

The STATE assumed by the parametric bias units will be encoded in free parameters and adapted together with the other free paramers of the policy network.

The method can be used to train a robot to produce a set of differentiated behaviors, such as walking forward, walking backward, and walking on the left or right side.



This project investigate whether functional specialization or modularity can support the development of multiple behaviors. With the term FUNCTIONAL specialization we refer to a situation in which one or more neurons, eventually forming a specific sub-part of the neural network policy, are primarily responsible for the production of a specific behavior and are less involved in the production of other behaviors. When the groups of specialized neurons are more connected with the neuron of the group than with other neurons, the groups of specialized neuron form a specialized neural module.

In principle, MODULAR solutions of this type can facilitate the development of multiple behavious since each module is responsible for the production of a different behavior. Consequently, the interferences that arise neural mechanisms supporting the production of different behaviors can be reduced.

The realization of a structural modularity of this type, however, also require the availability of regulatory mechanisms that enhance the activity of the neurons specialized for the production of the behavior that is relevant in the current context and suppress or filter-out the effect of the neurons specialized for the production of alternative behaviors.

REGULATION of the sort can be realized by using special neurons that control the desired impact of other neurons. For example, by including in the policy network standard neurons and logistic regulatory neurons and by multiplying the output of standard neurons for the activation of the associated regulatory neurons.

The project will thus involve the implementation of regulatory network of thus type and the realization of experiments involving the production of different behaviors.

4. Co-adaptation of swarms of competing agents

The goal of this project is to study the evolution of complex behaviour in self-play collective scenarios

This can be realized by using Magent, a tool that permits to simulate efficiently large groups of agents

More specifically the objective of this project is to identify methods that can promote the achievement of global progress

https://github.com/geek-ai/MAgent

The goal of this project is to study the evolution of complex behaviour in self-play collective scenarios. Self-play scenario consist of problems in which one or more adaptive agents adapt in an environment that include competing adaptive agents. Collective scenarios involve problem in which the adaptive agent is formed by a group of agents. This scenario is particularly interesting from the point of view of synthesizing complex behaviour since the adaptive agents are situated in an environment that can become progressively more difficult and since it expose the adaptive agents to highly variable condition.

THIS objective can be realized by using MAgent, i.e. a tool that permits to simulate efficiently hundreds to millions of agents on the GPU. In particular it can be realized by using the battle example included in the tool that involves two competing armies composed of hundreds of soldiers.

MORE specifically the objective of this project is to identify methods that can promote the achievement of global progress, i.e. the production of solutions that perform better and better against competitors of any kind and not only against the competitors encountered recently

In particular the project can verify the importance of evaluating the adapting

army against well-differentiated strategies selected among previous version of the learning army itself. Or the importance of filtering out opportunistic variations, i.e. variations that produce local progress instead than global progress.

The problem clearly requires to familiarize with the MAgent tool and requires the availability of suitable GPU

THE project investigate the evolution of cooperative and specialization skills in teams of air hockey players.

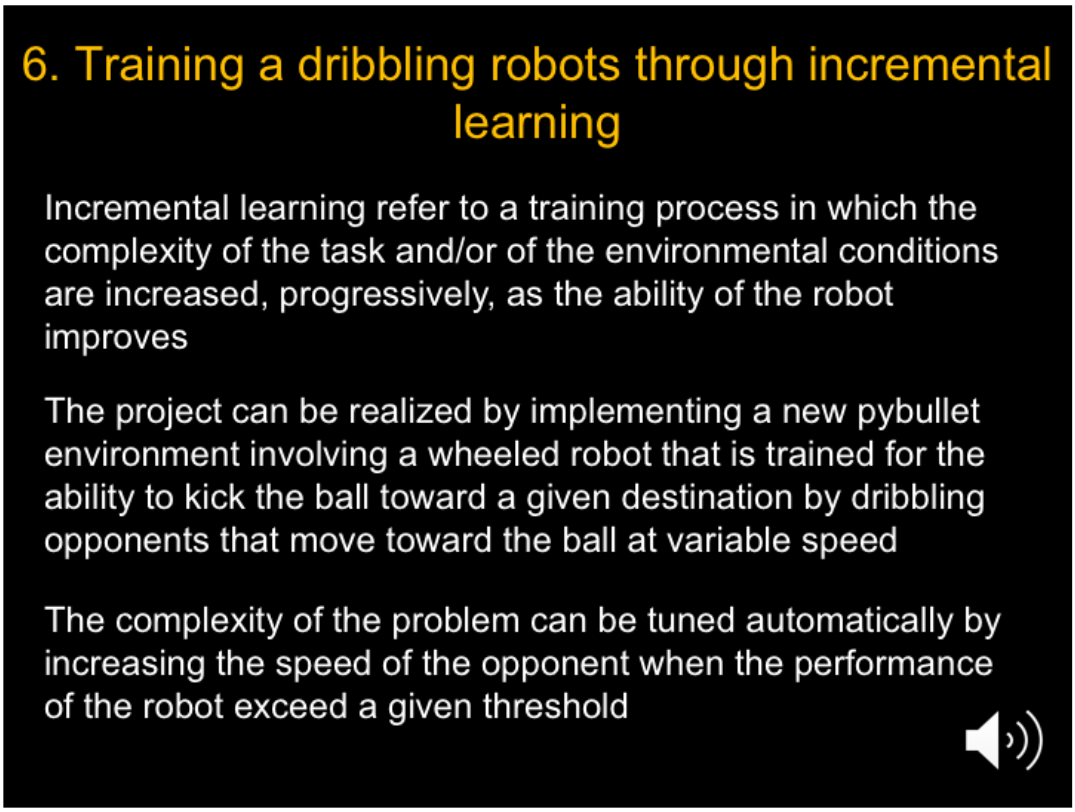
COOPERATION and specialization refer to the ability to cooperate with the companion to maximize the efficacy of the team behavior, the ability to avoid interfering with the companion, and the ability to specialize so to play complementary roles. The study will compare the results obtained by using homogeneous and heterogeneous teams, i.e. teams formed by individuals which operate on the basis of policy network that have identical or different parameters.

THE STUDY can be conducted by implementing a fast custom simulator of the game.

Players might receive as input:

a value that indicate whether are located on the left or right side of the table, and pre-elaborated visual information encoding their position and velocity with respect to the field, the angle and the distance of the disk, the angle and the distance of the companion, the angle and the distance of the two opponents.

The output of the players can control the acceleration of the player's paddle over the plane.



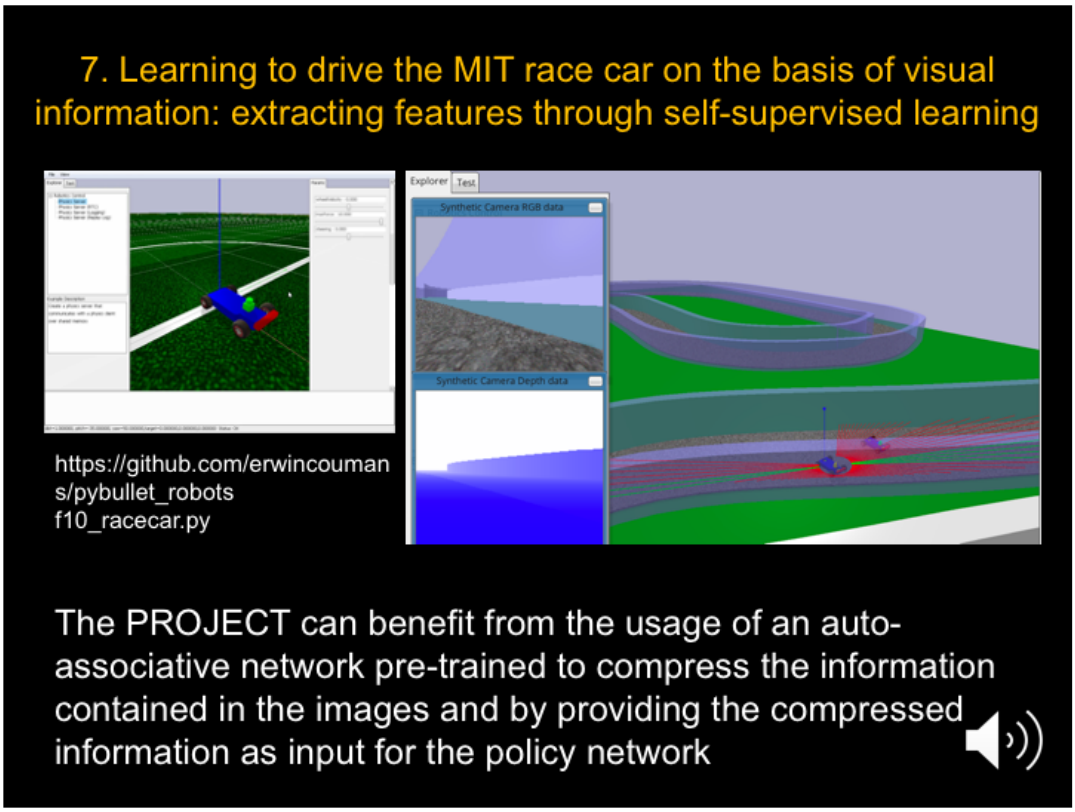
This project investigate the usage of incremental learning to train a dribbling robot.

INCREMENTAL learning refers to a training process in which the complexity of the task and/or of the environmental conditions are increased progressively as the ability of the robot improves.

THE PROJECT can be realized by implementing a new pybullet environment involving a wheeled robot that is trained for the ability to kick a ball toward a given destination by dribbling opponents that move toward the ball at variable speed.

THE COMPLEXITY of the problem can be tuned automatically by increasing the speed of the opponent when the performance of the robot exceed a given threshold.

The robot can be provided with an omnidirectional camera from which the robot can extract the position and the distance of the target, of the ball and of the opponent.



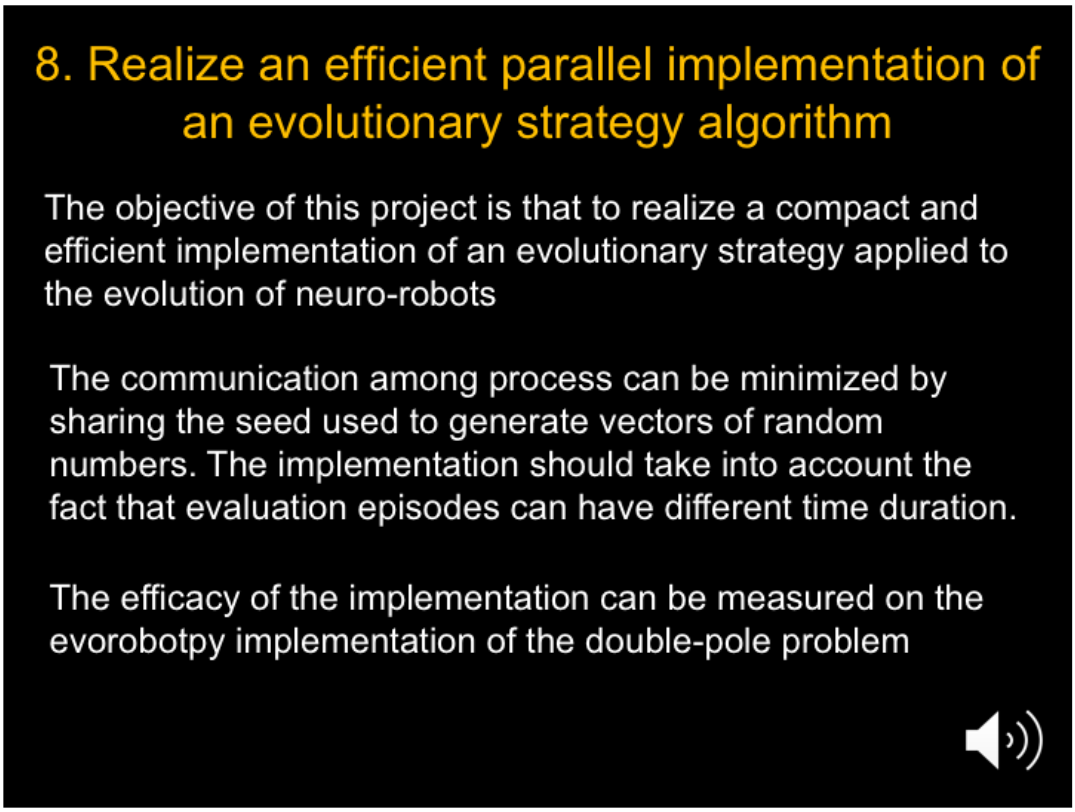
The objective of this project consist in training a racecar to drive in a simulated race track on the basis of visual information and to use this problem to study the combination of trial and error and unsupervised learning methods

This can be realized by creating an environmental problem based on a simuation of the MIT racecar. This video on the left shows the car and the effect of the actuators. The figure on the right shows a race track implemented in the f10_racecar.py script and the RGB and depth image perceived by the camera mounted on the car.

The training can be done by using evolutionary or reinforcement learning.

The PROJECT can benefit from the combined usage of trial and error learning and unsupervised learning. Unsupervised learning can be used to extract useful perceptual features. In particular, one can use an auto-associative network to compress the information contained in the images into an internal vector that is then used to recreate the original image. The features contained in the internal vectors can then be passed as input to the policy network (see for example (see D.Ha and J Schmidhuber (2018). World models, arXiv:1803.10122).

The training of the auto-associative network can also be described as a form of self-supervised learning since it relies on a supervised learning algorithm that requires the specification of the teaching input but exploits the fact that the teaching input is provided directly by the environment without the intervention of the experimenter.



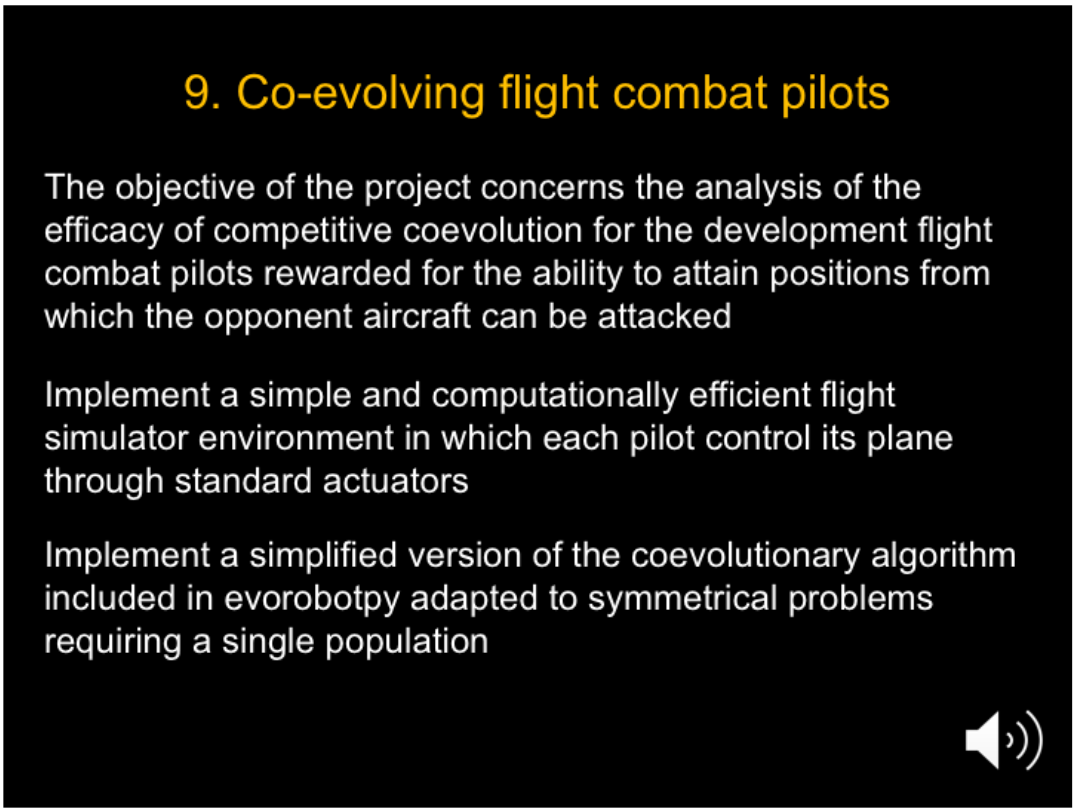
An advantage of evolutionary strategies is that they can be parallelized so to obtain an almost linear gain in speed (see Salimans T., Ho J., Chen X., Sidor S & Sutskever I. (2017). Evolution strategies as a scalable alternative to reinforcement learning. arXiv:1703.03864v2).

THE objective of this project is that to realize a compact and efficient implementation of an evolutionary strategy applied to the evolution of neuro-robots.

The process can be parallelized at the level of the popupulation and eventually at the level of the evaluation episodes through the usage of a multiprocessor library such as mpi4py.

THE communication among process can be minimized by sharing the seed used to generate vectors of random numbers. Moreover, the implementation should take into account the fact that evaluation episodes can have different time duration.

THE efficacy of the implementation can be measured on the evorobotpy implementation of the double-pole problem by measuring the gain with variable number of thread/processors.



This other project involves the co-evolution of flight combact pilots.

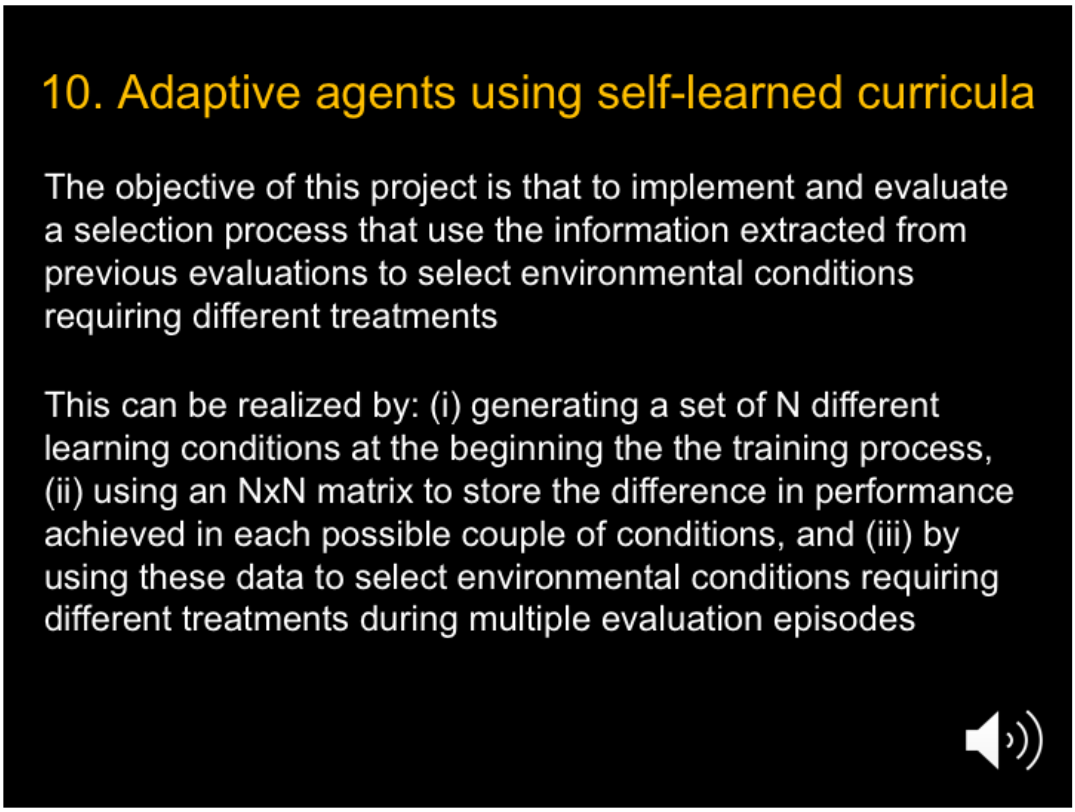
Competitive co-evolution is an ideal framework for developing general, robust, and high-performing solutions. This since the need to cope with adapting competitors maintain a constant selection pressure for the development of innovations and expose adaptive robots to highly varied learning experiences.

THE objective of the project concerns the analysis of the efficacy of this technique in the context of the evolution of flight combat pilots rewarded for the ability to win combact games.

THIS can be realized by implementing a simple and computationally efficient flight simulator environment in which each pilot control its plane through standard actuators (e.g. throttle, elevator, rudder). The goal of the pilot is that to attain positions from which the opponent aircraft can be attacked and to avoid positions from which it can be attacked. Pilots observation will include the relative position, orientation, and distance of the opponent aircraft.

THE experiments can be conducted by using a simplified version of the co-

evolutionary algorithm included in evorobotpy software. The original algorithm is designed for asymmetrical problems like predator and prey robots that require to evolve two separate populations and should thus be simplified to be applied to a symmetrical problem, like flight combact, that requires a single population.



This project address the possibility to automatically select leaning experiences that facilitate the development of general solutions.

The outcome of an adaptive process is strongly influenced by the conditions experienced during robots' evaluation. In standard approach this problems is normally approached by selecting the initial environmental conditions randomly. This method is simple and can support the development of solutions that are robust with respect to environmental variations.

However, when unfrequent conditions required qualitatively different treatments than more frequent conditions, the adaptive process might converege on solutions optimized for frequent conditions only.

THE OBJECTIVE of this project is that to implement and evaluate a selection process that use the information extracted from previous evaluations to select environmental conditions requiring different treatments.

THIS CAN be realized by: (i) generating a set of N different learning conditions at the beginnig the the training process, (ii) using an NxN matrix to store the difference in performance achieved in each possible couple of

conditions, and (iii) by using these data to select environmental conditions requiring different treatments during multiple evaluation episodes.

The matrix can be updated on the basis of the performance achieved by the top individuals of each generation over multiple evaluation episodes.

Environmental conditions equiring different treatments can be chosen by selecting the first condition randomly, and by using the matrix to identify conditions that produced different result from the already chosen conditions

The data contained in the matrix can can also be used to select conditions that have an intermediate level of difficulty, on the average. In other words conditions that are neither too easy neither too difficult at the current stage of the evolutionary process.

The approach can be validated for example on the double pole balancing problem.

11. Create your locomotor environment based on an real robotic platform

This goal of this project is that to implement a locomotor environment based on an real biped or quadruped robot

Laikago robot

The environment should be validated and by training the robot through evolutionary or reinforcement learning algorithm

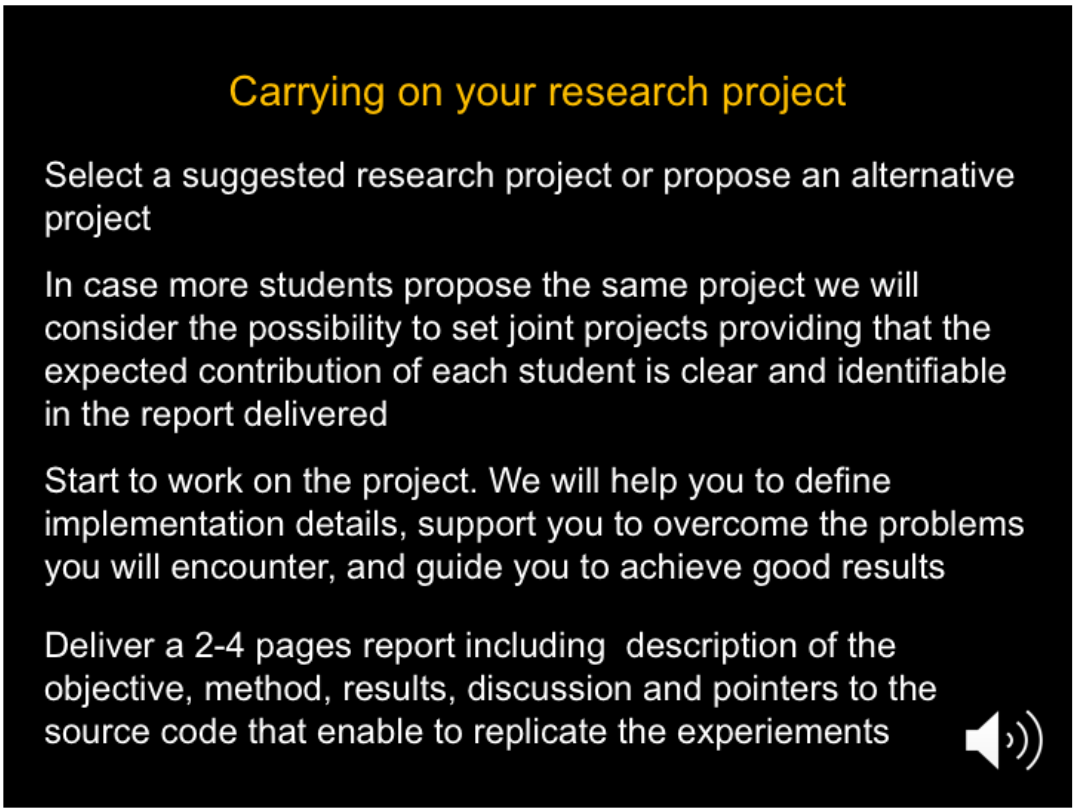
The results collected by training the robot should be used to refine the environment

FINALLY, this project concerns the development of a locomotor environment based on an real biped or quadruped robotic platform such as the Atlas humanoid robot or the Laikago robot shown in this slide.

As I explained in the intro to exercise 5, the universal robot description format files for several real robotic platform are already available in pybullet. What you should implement, to carry on this project, are the gym methods that permit to reset an episode and perform an evaluation step. Moreover, you should decide the information to be encoded in the observation, the actuators to be controlled by the output of the network policy, and the reward function.

THE environment should be validated and refined by training the robot through evolutionary or reinforcement learning algorithm.

THE results collected by training the robot should be used to refine the environment, that is to vary the design choice in a way that maximize the possibility to synthesize effective solutions



The research project will constitute the most important element for your evaluation. So please select it and start to work on it as soon as possible.

We can proceed in the following way.

FIRSTABLE you are asked to select and communicate to me and Vladislav one of the suggested research project. If you like, you can propose an alternative research project that we will evaluate.

IN CASE more students propose themselves for the same project we will consider the possibility to set joint projects providing that the expected contribution of each student is clear and will be specified in the report delivered.

ONCE projects are assigned, we will discuss with each of your the details of how the project can be realized. The description of the projects that I made in this presentation is very synthetic. So do not worry if you do not yet have a clear picture of how to realize the project that you would like to select. I will provide more details and additional explanations to you before you start to work on your project and I will be available to provide support and help while you work on the project

Finally, once you collected your results, you should submit a short 2-4 pages report, organized as a scientific paper with a description of the objective, method, results, and a brief discussion. The report should also point to the source code developed.

Research project are by definition risky. Consequently, the evaluation will be based on the work done and on the comprehension of the theoretical aspects associated to the project. The outcome of the evaluation will not be influenced by the scientific significance of the results obtained.

Those of you that will manage to obtain scientifically significant results can consider, after the end of the course, the possibility to extend the report in a scientific paper that can be submitted to a conference or to a journal. In case you decide to do that, I will be available to contribute to the preparation of the article.